



An Introduction to Some Fundamental Aspects of EXAFS Analysis: Fourier Concepts and Random Errors

Corwin H. Booth¹ and Yung-Jin Hu^{2,3}

¹Chemical Sciences Division
Glenn T. Seaborg Center

Lawrence Berkeley National Laboratory
²Nuclear Sciences Division

Lawrence Berkeley National Laboratory
³Department of Chemistry

University of California, Berkeley

Presented at the SSRL Workshop: Advanced Topics in EXAFS Analysis and Applications, Oct. 15, 2008
LAWRENCE BERKELEY NATIONAL LABORATORY

Why worry about random noise in EXAFS?

- Aren't the errors in EXAFS analysis primarily due to systematic problems?
- Random error motivations
 - must understand whether in a random limit, or how far away
 - If so (or close), can use standard statistical concepts
 - Even if not close, understanding effects of random errors gives a deeper understanding into the ultimate power and limitations of the technique
- *This talk is intentionally pedantic!*

LAWRENCE BERKELEY NATIONAL LABORATORY

Topics

- Basic concepts:
 - Fourier rules: data sampling, Nyquist, data padding
 - information content: independent data, resolution
- Random noise:
 - Fun with χ^2
 - parameter p_i , error determination
 - determination of number of independent data points
 - Rough effect of systematic error
 - The F-test: fit-model significance
- Example of random-noise limited system
- Example of systematic-noise limited system

LAWRENCE BERKELEY NATIONAL LABORATORY

Prelims: Fourier concepts

- Data set description:
 - number of data points in k -space N_k
 - sampling interval δ_k
 - maximum k is $k_{\max} = N_k \delta_k + 1$
 - data are fit over some range Δ_k and Δ_r
 - conventional k -space fit (no Fourier filtering) every data point is independent (assuming good resolution), so

$$N = N_k \Delta_k / k_{\max}$$
- Fourier transforming indicates:
 - highest "frequency" $r_{\max} = \pi/(2\delta_k)^{-1}$ (Nyquist frequency $f_N = 2r_{\max}$)
e.g. for sampling interval $\delta_k = 0.05 \text{ \AA}^{-1}$, $r_{\max} = 31 \text{ \AA}$
 - for N_k , discrete Fourier transform has N_k , too! But complex (real and imaginary), therefore resolution is half FT interval and

$$\delta_R = 1/2 r_{\max}/(N_k/2) = \pi/[4\delta_k k_{\max}/(2\delta_k)] = \pi/(2k_{\max})$$
,
e.g. $k_{\max} = 15 \text{ \AA}^{-1}$, $\delta_k = 0.1 \text{ \AA}$

LAWRENCE BERKELEY NATIONAL LABORATORY

More Fourier concepts: Independent data points

- Spectral theory indicates that each point in k -space affects every point in r -space. Therefore, assuming a fit range over k (and r):
- $$\Delta_r = r_2 - r_1 \quad \Delta_k = k_2 - k_1$$
- $$r_{\max} = \frac{\pi}{2\delta_k} \quad N_k = \frac{k_{\max}}{\delta_k} = \frac{r_{\max}}{\delta_r}$$
- $$N_{\text{ind}} = N_k \frac{\Delta_r}{r_{\max}} \frac{\Delta_k}{k_{\max}} = \frac{2}{\pi} \Delta_r \Delta_k \quad \text{Fourier result}$$
- $$N_{\text{ind}} = \frac{2}{\pi} \Delta_r \Delta_k + 2 \quad \text{"Stern's rule" EXAFS result}$$
- Stern, Phys. Rev. B 48, 9825 (1993).
- Fit degrees of freedom $v = N_{\text{ind}} - N_{\text{fit}}$
 - Generally should never have $N_{\text{ind}} > N_{\text{fit}}$ ($v < 1$)
 - But what does this mean? It means that:

$$\text{For } N_{\text{fit}} \text{ exceeding } N_{\text{ind}}, \text{ there are other linear combinations of } N_{\text{fit}} \text{ that produce EXACTLY the same fit function}$$

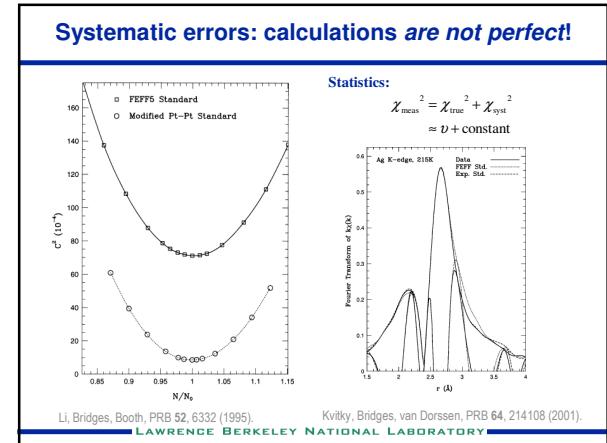
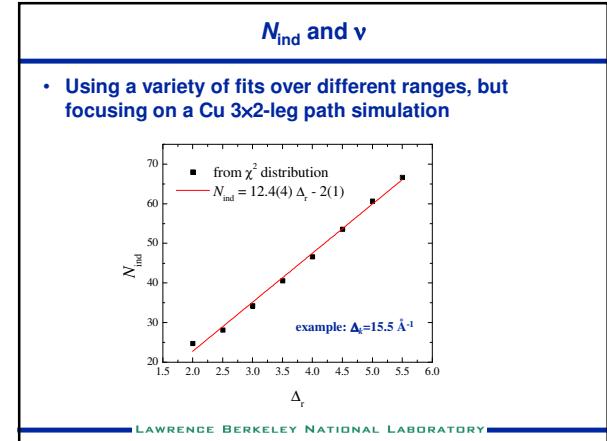
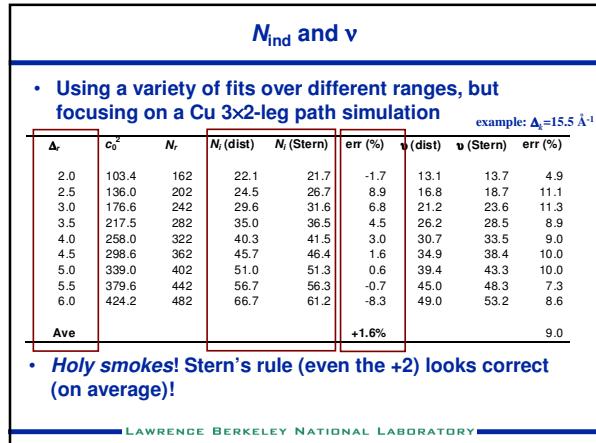
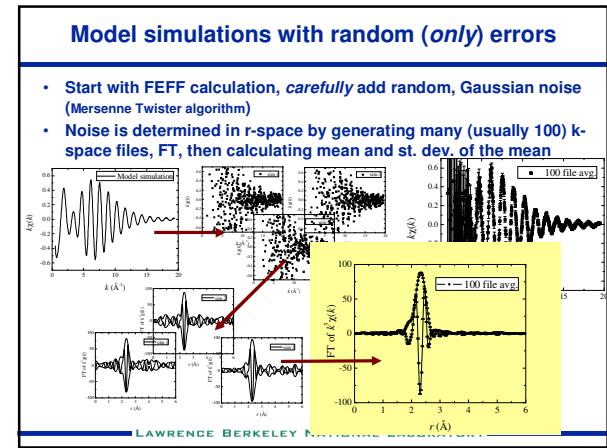
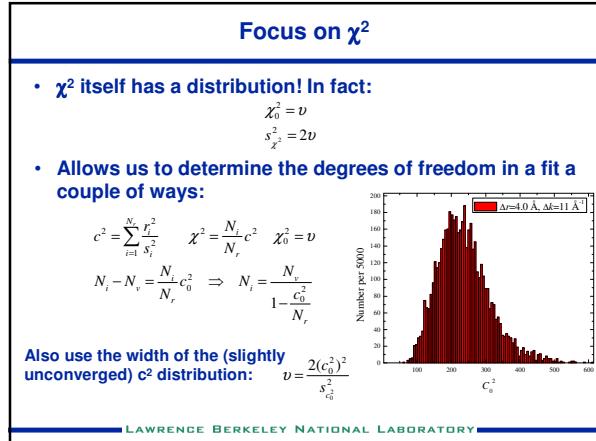
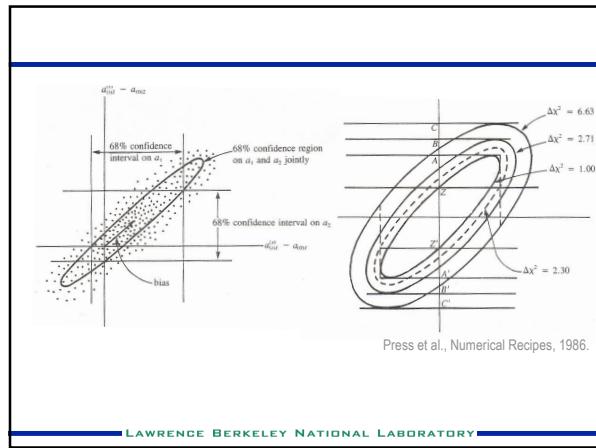
LAWRENCE BERKELEY NATIONAL LABORATORY

Assuming randomly distributed errors

- Conventional error analysis revolves around the statistical χ^2 :

$$\chi^2 = \sum_{i=1}^{N_{\text{ind}}} \frac{(y_i - \bar{y}_i)^2}{s_i^2} = \sum_{i=1}^{N_{\text{ind}}} \frac{r_i^2}{s_i^2}$$
- Two problems:
 - Must sum only over independent data points N_{ind} , or normalize by a factor N/N_r
 - Must know the errors s_i (variance = s_i^2)
- Assuming you know these things, can determine best fit $\langle y \rangle$ using a model with some parameter set p_1, p_2, \dots
- Errors in p_i , s_i can be determined by refitting with fixed $p_i = \langle p \rangle \pm s_{p_i}$, at which point $\chi^2 = \chi_0^2 + 1$

LAWRENCE BERKELEY NATIONAL LABORATORY



Statistical tests between models: the F-test

$F = \frac{\chi^2 / v_1}{\chi^2 / v_0}$ $F = \frac{(\chi^2 - \chi^2_0)/(v_i - v_0)}{\chi^2_0 / v_0}$ $R \propto \sqrt{\sum_i (y_i - \bar{y}_i)^2}$

$$F = \left[\left(\frac{R_1}{R_0} \right)^2 - 1 \right] \frac{(n-m)}{b}$$

$$\alpha = P(F > F_{b,n-m,\alpha}) = 1 - I_{\chi^2} \left[\frac{n-m}{2}, \frac{b}{2} \right]$$

where α is the confidence level and $I_x[y,z]$ is an incomplete beta function

- F-tests have big advantages for data sets with poorly defined noise levels**
 - with systematic error, R_1 and R_0 increase by ~constant, reducing the ratio (right direction!)

LAWRENCE BERKELEY NATIONAL LABORATORY

F-test example testing resolution in EXAFS: knowledge of line shape huge advantage

• 6 Cu-Cu at 2.55 Å, 6 at 2.65 Å
 • $k_{\max} = \pi/(2\delta_R) = 15.7 \text{ \AA}^{-1}$
 • 1p (w/ cumulants) vs 2p fits
 • Fits to 18 Å⁻¹ all pass the F-test
 • $S_n=50$ does not pass at 15 Å⁻¹
 • $S_n=20$ does not pass at 13 Å⁻¹
 • $S_n=4$ does not pass at ~11 Å⁻¹

LAWRENCE BERKELEY NATIONAL LABORATORY

Unfortunately, systematic errors dominate

Example without systematic error:
 $k_{\max}=13 \text{ \AA}^{-1}, v_0=6.19, v_1=7.19$
 $R_0=0.31$
 $R_1=0.79$
 $\alpha=0.999 \rightarrow \text{passes F-test}$

Example with systematic error at $R=2\%$:
 $k_{\max}=13 \text{ \AA}^{-1}, v_0=6.19, v_1=7.19$
 $R_0=0.31+2.0=2.31$
 $R_1=0.79+2.0=2.79$
 $\alpha=0.86 \rightarrow \text{does not pass}$

LAWRENCE BERKELEY NATIONAL LABORATORY

Real world example #1: Plutonium on Manganese Oxide Model

Test model

- Pu-O₀₁:
 $R = 2.33 \text{ \AA}$
 $CN = 9$
- Pu-Mn:
 $R = 3.82 \text{ \AA}$
 $CN = 4$
- Pu-O₀₂:
 $R = 3.70 \text{ \AA}$
 $CN = 2$
- Pu-O₀₃:
 $R = 3.99 \text{ \AA}$
 $CN = 2$

- EXAFS measurements on environmental systems are commonly limited by random noise!
- Here, want to know local structure and redox properties of Pu sorbed onto pyrolusite (MnO_2)
- Environmental conditions (eg., potential repository failure) ~equivalent to 0.000005% to 0.0005% Pu on MnO_2
- Key question:** Is Pu coordinated to surface of the mineral? That is, can we see a Mn peak in the EXAFS?

LAWRENCE BERKELEY NATIONAL LABORATORY

Plutonium on Manganese Oxide

- 15 scans of Pu on Mn oxide (left)
- Average with standard deviation of the mean (right)
- MnO_2 : 0.05%Pu
- 1 scan took about 15 minute, total time ~ 4 hours
- SSRL, BL 11-2, <<125 kcps, T= 30 K, blah blah blah

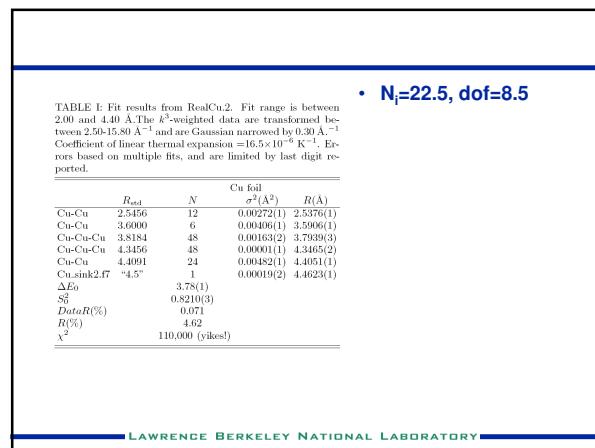
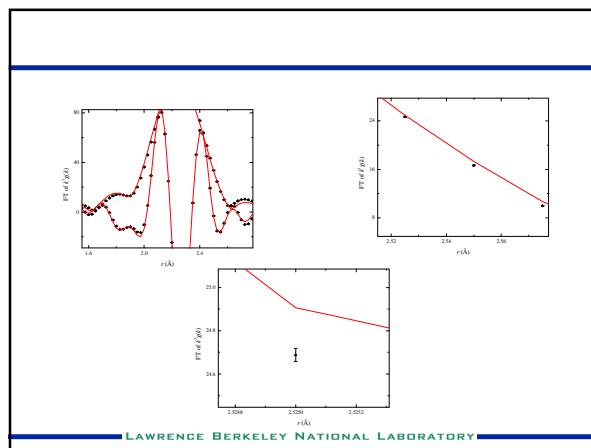
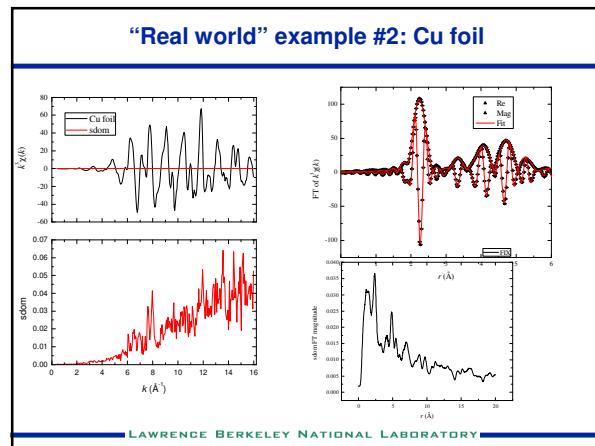
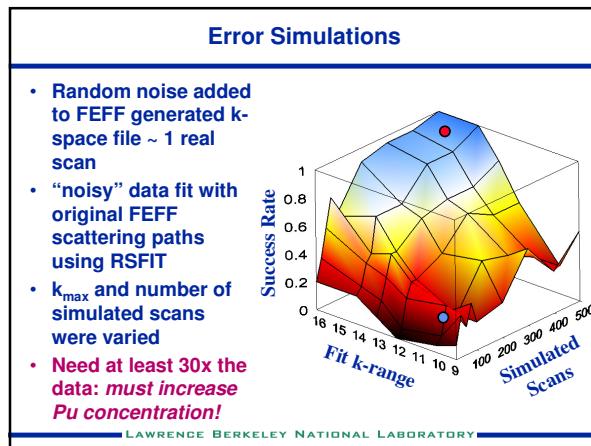
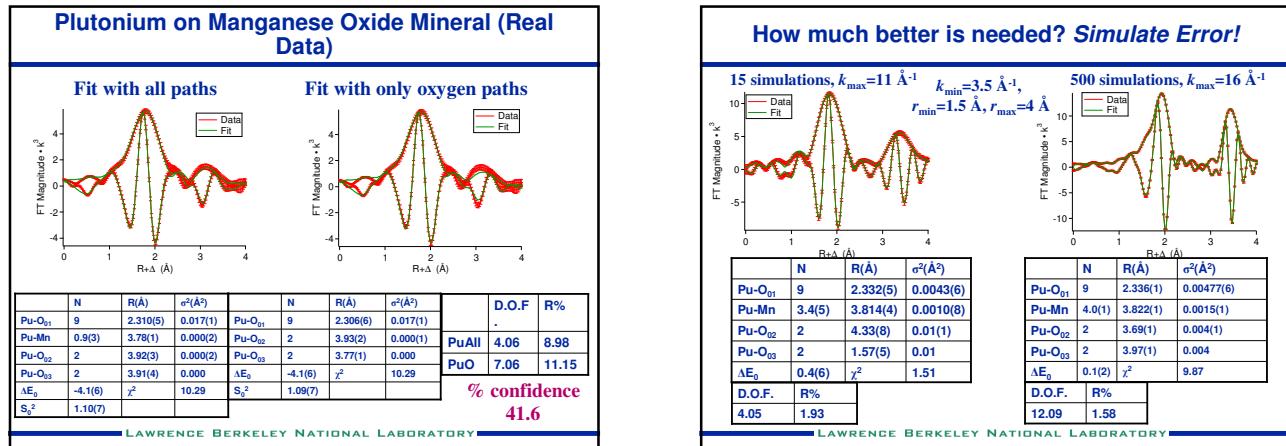
LAWRENCE BERKELEY NATIONAL LABORATORY

Plutonium on Manganese Oxide Mineral (Real Data)

Fit with all paths Fit with only O_{shell01} and Mn path

	N	R(Å)	$\sigma^2(\text{Å}^2)$		N	R(Å)	$\sigma^2(\text{Å}^2)$		D.O.F	R%
Pu-O ₀₁	9	2.310(5)	0.017(1)	Pu-O ₀₁	9	2.308(6)	0.017(1)	PuAll	4.06	8.98
Pu-Mn	0.9(3)	3.78(1)	0.000(2)	Pu-Mn	30(15)	3.99(1)	0.045(7)	PuMn	7.06	14.4
Pu-O ₀₂	2	3.92(3)	0.000(2)	ΔE_0	-3.8(7)	χ^2	25.84	% confidence		
Pu-O ₀₃	2	3.91(4)	0.0000	S_0^2	1.00(7)			77.1		
ΔE_0	-4.1(6)	χ^2	10.29							
S_0^2	1.10(7)									

LAWRENCE BERKELEY NATIONAL LABORATORY



Systematic vs. Random error

- Systematic errors for nearest-neighbor shells are about 0.005 \AA in R , ~5% in N , 10% in σ (Li, Bridges, Booth 1995)
- Systematic error sources:
 - sample problems (pin holes, glitches, etc.)
 - correction errors: self-absorption, dead time, etc.
 - backscattering amplitudes
 - overfitting (too many peaks, strong correlations between parameters)
- Random error sources:
 - some sample problems (roughly, small sample and moving beam)
 - low counts (dilute samples)

LAWRENCE BERKELEY NATIONAL LABORATORY

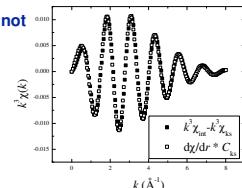
Issues with DW: #1 k vs r space DW

- The standard k-space form of the Debye-Waller term is very good, but not exact:

$$\int_{-\infty}^{\infty} e^{-q^2 r^2} \sin[p(x+\lambda)] dr = \frac{\sqrt{\pi}}{q} e^{-\frac{p^2}{4q^2}} \sin(p\lambda)$$

$$\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(r-R)^2}{2\sigma^2}} \sin(2kr) dr = e^{-2k^2\sigma^2} \sin(2kR)$$

$$\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(r-R)^2}{2\sigma^2}} \frac{e^{\frac{2x}{\lambda}}}{r^2} \sin(2kr) dr \approx e^{-2k^2\sigma^2} e^{-\frac{2R}{\lambda}} \frac{\sin[2k(R - \frac{2\sigma^2}{R}(1 + \frac{R}{\lambda}))]}{R^2}$$

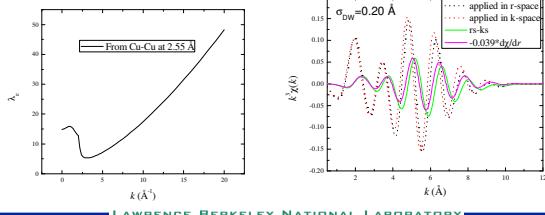


Crozier and Seary, Can. J. Phys., 58 1388 (1980).
G. Bunker, Nucl. Inst. Methods, 207, 437 (1983).
Eisenberger and Brown, SSC 29, 481 (1979).

LAWRENCE BERKELEY NATIONAL LABORATORY

Issues with DW: #1 k vs r space DW

- There is a negative phase shift in the EXAFS equation as a consequence of the $\exp(-2r/\lambda_e)/r^2$ term:
 $\nabla \rightarrow -4k\sigma^2(1+R/\lambda_e)/R$ (FEFF doesn't do this)
- $\lambda_e \sim 10 \text{ \AA}$, so this acts like a negative R -shift of $2\sigma^2(1+R)/R$!
- For $R=2.55 \text{ \AA}$ and $\sigma^2=0.04 \text{ \AA}^2$, $\Delta R_{\text{meas}}=-0.04 \text{ \AA}$
- BUT, two probs: $\lambda_e(k)$ and $F(k)$ are r -dependent!



LAWRENCE BERKELEY NATIONAL LABORATORY

Conclusion

- Understanding basic principles of error analysis is useful, even in some practical cases
- Understanding random errors highlights the still strong role that systematic errors (lineshape, in $F(k)$) have

Acknowledgements



LAWRENCE BERKELEY NATIONAL LABORATORY